CLEARINGHOUSE FOR FEOERAL SCIENTIFIC ANO TECHNICAL INFORMATION CFSTI
OOCUMENT MANAGEMENT BRANCH 410.11

LIMITATIONS IN REPROOUCTION QUALITY

ACCESSION # AD 603989

☒ 1. WE REGRET THAT LEGIBILITY OF THIS DOCUMENT IS IN PART UNSATISFACTORY. REPRODUCTION HAS BEEN MAOE FROM BEST AVAILABLE COPY.

☐ 2. A PORTION OF THE ORIGINAL DOCUMENT CONTAINS FINE DETAIL WHICH MAY MAKE READING OF PHOTOCOPY OIFFICULT.

☐ 3. THE ORIGINAL DOCUMENT CONTAINS COLOR, BUT OISTRIBUTION COPIES ARE AVAILABLE IN BLACK-AND-WHITE REPROOUCTION ONLY.

☐ 4. THE INITIAL OISTRIBUTION COPIES CONTAIN COLOR WHICH WILL BE SHOWN IN BLACK-AND-WHITE WHEN IT IS NECESSARY TO REPRINT.

☐ 5. LIMITEO SUPPLY ON HANO: WHEN EXHAUSTEO, OOCUMENT WILL BE AVAILABLE IN MICROFICHE ONLY.

☐ 6. LIMITEO SUPPLY ON HANO: WHEN EXHAUSTED DOCUMENT WILL NOT BE AVAILABLE.

☐ 7. OOCUMENT IS AVAILABLE IN MICROFICHE ONLY.

☐ 8. OOCUMENT AVAILABLE ON LOAN FROM CFSTI (TT OOCUMENTS ONLY).

☐ 9.

NBS 9/64

PROCESSOR: RG

COPY 1 of 1 COPIES

COMMENTS ON H. J. BARNETT'S
"SPECIFIC INDUSTRY OUTPUT PROJECTIONS"

A. W. Marshall

P-243

1 October 1951

The RAND Corporation
SANTA MONICA · CALIFORNIA

COMMENTS ON H. J. BARNETT'S "SPECIFIC INDUSTRY OUTPUT PROJECTIONS"

A. W. Marshall

Two comments come to mind with regard to Dr. Barnett's paper, "Specific Industry Output Projections." The first of these concerns the appropriateness of comparisons of the kind carried out in the paper for reaching decisions as to the acceptance or rejection of proposed forecasting methods; the second comment concerns an alternative method of measuring the forecasting errors.

It would appear, given the comparisons in Tables 3 and 4, based upon the mean deviations of actual from predicted industry outputs, that projection methods using input-output tables are not very much better than quite elementary "naive" model methods. Indeed the multiple regression forecasts seem to be somewhat better than those based upon input-output relations. Where "naive" models have in some sense to be taken seriously; e.g., if asked to forecast output by industry, for 1956 say, I would prefer to use Barnett's multiple regression method rather than the input-output method exhibited here; it is well to keep in mind their purpose and character. They are not intended to be legitimate alternatives to the model or procedure being tested, but rather are designedly crude and inefficient, almost <u>reductio ad absurdum</u> constructions of economic models and forecasting procedures. They represent a level of efficiency so low and so easily attained that any forecasting procedure seriously proposed for operational use, which cannot almost uniformly do better than they can, must be rejected as being operationally unacceptable. Two warnings are needed here. First, Barnett's multiple regression model must be conceded to be

"semi-naive" in the sense that even if we were to add additional variables to the equations which were thought to have a special relevance for the explanation of the output of some specific industry, it is unlikely, due to the correlation between most economic time series, that continued large reductions in the sum of squares about the regression line can be obtained. In other words the substitution and/or addition of other explanatory variables will probably not greatly increase the forecasting efficiency of these regression equations; we are already, very likely, in an area of diminishing returns to scale in this sense. Second, the kinds of comparisons made in Barnett's paper are very appropriate to decisions as to whether or not a certain method of forecasting should or should not be used in practice, given its current stage of development, but are often of minor importance with regard to decisions concerning the advisability of continuing development of these methods. Thus this type of competitive trial of serious, though perhaps immature, models and forecasting methods against "naive" models should not lead anyone to discard, or neglect the development of, really promising techniques.

As a second comment I should like to suggest an alternative and somewhat more natural measure, at least to a mathematician, of the error of prediction of the various methods of projecting or estimating specific industry outputs in some future year. In Tables 3 and 4 Barnett has used as his measure of error $\sum_{i=1}^{28} |X_i(GNP^*) - X_i(A)|$, where X_i (GNP*) denotes the estimated output, in terms of either an index number or dollar value, of the ith industry based upon the estimate GNP* of GNP, and $X_i(A)$ denotes the actual output of the ith industry. All of the above, of course, refers to some fixed year and method of forecasting. As an alternative it

is appealing to think of the observed production by industry and the projected productions as vectors in n dimensional Euclidean space and to think of the error of the projection as being the distance between the two points. Each vector has 28 components in Barnett's case and the distance between the two points (vectors) is given by

$$d = \left(\sum_{i=1}^{28} \left\{ X_i(\mathrm{GNP}*) - X_i(A) \right\}^2 \right)^{\frac{1}{2}} .$$

Not only is this the more usual definition of the distance between two vectors but it is also a measure which fits in with what would seem to be, from the statistical point of view, the aim of research in forecasting methods, i.e., the finding of minimum variance estimates of the future values of economic variables. From this point of view we also see that once it is decided what it is we wish to forecast, all questions of further disaggregation resolve themselves into questions as to whether or not a particular disaggregation reduces the variance of our forecasts.

One additional comment is to be made. Since it seems to be almost certain that in the future we will have estimates of GNP which have considerably smaller errors than the estimates used in the present paper, whose accuracy was compromised by secondary consumption or investment considerations, some separation of the total error of the various forecasting techniques into its component parts is desirable. Errors of the order of magnitude made in the GNP estimates are so large that in some sense none of the methods obtained a fair trial. It is, of course, in general desirable to be able to factor out the errors contributed by the separate steps in the forecasting methods since one method may be much more sensitive than

another to errors in some common component, say the first component in all of the forecasting methods in Barnett's paper, the estimate of GNP. This factorization is easily carried out in principle as may be seen as follows: Let us denote by $X_i(GNP)$ the estimated output of the ith industry we would have made if we had known the true value of GNP; then we have

$$d_1 = \left(\sum_{i=1}^{28} \left\{ X_i(GNP - X_i\ (A) \right\}^2 \right)^{\frac{1}{2}}$$

as a measure of the error we would have made even if we had had the best possible knowledge of the value of GNP. Also, we have

$$d_2 = \left(\sum_{i=1}^{28} \left\{ X_i(GNP^*) - X_i(GNP) \right\}^2 \right)^{\frac{1}{2}} ,$$

and thus the total error d is separated into two factors of which it is the vector resultant. As an example I have performed this factorization for Barnett's multiple regression method where one can easily obtain the $X_i(GNP)$ from the equations

$$X_i(GNP) = a_i + b_i\ GNP + c_i t$$

by substituting the correct value of GNP (152,400) rather than the estimate GNP* used to obtain the values in Tables 3 and 4. A similar factorization for the input-output method would, of course, entail much more work. Working with the consumption model (GNP* = 170,000) and the dollar value figures in millions of 1939 dollars we obtain the results listed in Table 1, below, along with the comparable figure of d for the input-output method.

TABLE I

COMPONENTS OF ERROR

[illegible]	Multiple Regression	Input-Output
d (total errors)	5,325	7,178
d_1 (error given exact GNP estimate)	3,354	-----
d_2 (GNP* component of error)	6,052	-----

Thus using an estimate of GNP which is too large by 11.5% leads to an overall increase in the error of forecast of 59.0%. The reader will also notice that the distance between the two estimates, one based upon GNP and the other upon GNP* are farther from one another than each estimate is from the true value. If we were concerned with a two-industry world this situation would be illustrated as in Chart 1.

CHART 1